PARAMETRIZATION OF KLOOSTERMAN SETS AND SL_3-KLOOSTERMAN SUMS

EREN MEHMET KIRAL[1], MAKI NAKASUJI[2]

1. Introduction

We give explicit and comprehensible formulas for the SL_3 long word Kloosterman sum, and related mathematical objects. This proceeding is a survey of the results in [KN20].

Our work is motivated by aesthetic considerations, believing that a beautiful expression for a Kloosterman sum would increase its comprehensibility and its recognizability when encountered elsewhere in nature. We hope that this work encourages the use of the explicit form of the Kloosterman sum, and leads to deeper results, better bounds and discovery of new identities for moments of L-functions.

1.1. **Definitions.** The generalized Kloosterman sums are defined as certain exponential sums on $\mathrm{U}_r(\mathbb{Z})$–double-cosets on matrix groups $\mathrm{SL}_r(\mathbb{Z})$. Here we denote the group of $r \times r$ unipotent matrices, i.e. upper triangular matrices with 1's on the diagonal entries, by U_r. We will drop the r from the notation when we fix its value throughout a section.

For a vector $\mathbf{c} \in (\mathbb{R}^*)^{r-1}$ define, $t(\mathbf{c}) := \mathrm{diag}(c_1, c_2/c_1, c_3/c_2, \ldots, 1/c_{r-1})$ and $w \in W$ a Weyl group element of SL_r, define

$$\Omega_w(\mathbf{c}) := \{u_L w t(\mathbf{c}) u_R \in \mathrm{SL}_r(\mathbb{Z}) : u_L, u_R \in \mathrm{U}_r\}.$$

For a matrix $A \in \mathrm{SL}_r(\mathbb{Z}) \cap BwB$ the c_i are integers given by minors of A, and they not changed upon multiplication by elements of U from either side. Therefore we obtain the stratification

$$\mathrm{U}(\mathbb{Z})\backslash BwB \cap \mathrm{SL}_r(\mathbb{Z})/\,\mathrm{U}(\mathbb{Z}) = \bigcup_{\mathbf{c}\in\mathbb{Z}^{r-1}} \mathrm{U}(\mathbb{Z})\backslash\Omega_w(\mathbf{c})/\,\mathrm{U}(\mathbb{Z}).$$

of the double $\mathrm{U}(\mathbb{Z})$-coset Bruhat cell into finite sets indexed by integral lattice points.

Let $\mathbf{n} = (n_1, n_2, \ldots, n_{r-1}) \in \mathbb{Z}^{r-1}$ and define the additive character $\psi_{\mathbf{n}}$ as follows: Let u be a unipotent matrix, where for $i < j$ its entries are denoted by $u_{i,j}$. Then

$$\psi_{\mathbf{n}}(u) = e(n_1 u_{1,2} + n_2 u_{2,3} + \cdots + n_{r-1} u_{r-1,r}).$$

For $\mathbf{m}, \mathbf{n} \in \mathbb{Z}^{r-1}$ the usual Kloosterman sum is defined as

$$S_w(\mathbf{m}, \mathbf{n}; \mathbf{c}) := \sum_{\substack{A\in \mathrm{U}(\mathbb{Z})\backslash\Omega_w(\mathbf{c})/\,\mathrm{U}(\mathbb{Z})\\ A=u_L w t(\mathbf{c}) u_R}} \psi_{\mathbf{m}}(u_L)\psi_{\mathbf{n}}(u_R). \tag{1.1}$$

2010 *Mathematics Subject Classification.* 11L05 (primary), and 11F22, 11F72 (secondary).

Key words and phrases. Kloosterman sums, Bott-Samelson factorization, Weyl Group, reduced word decomposition, Bruggeman-Kuznetsov trace formula.

1. Supported by Grant-in-Aid for JSPS Research Fellow 18F18326.
2. Supported by Grant-in-Aid for Scientific Research (C) 18K03223.

This sum is well defined if $\mathbf{m}$ and $\mathbf{n}$ satisfy certain conditions.

Specifically, the Kloosterman sets of the "*big cell*" in SL_3 are written as

$$\Omega_{w_0}(c_1, c_2) = \left\{ A \in \mathrm{SL}_3(\mathbb{Z}) : A \in \mathrm{U}_3(\mathbb{Z}) w_0 \begin{pmatrix} c_1 & & \\ & \frac{c_2}{c_1} & \\ & & \frac{1}{c_2} \end{pmatrix} \mathrm{U}_3(\mathbb{Z}) \right\}, \tag{1.2}$$

where c_1, c_2 are nonzero integers and the set of conditions on $\mathbf{m}$ and $\mathbf{n}$ that need to be satisfied is void. The long word SL_3 Kloosterman sum with modulus $\mathbf{c} = (c_1, c_2)$ can be described as a sum over $\Omega_{w_0}(c_1, c_2)$. In this paper we give a finer decomposition of (1.2) via the sets $\Omega(d_1, d_2, f)$ defined as follows. Given d_1, d_2, f nonzero integers, define,

$$\Omega(d_1, d_2, f) := \left\{ A \in \mathrm{SL}_3(\mathbb{Z}) \,|\, \gcd(A_{31}, A_{32}) = f, A_{31} = d_1 f, M_{\{23\},\{12\}} = d_2 f \right\}. \tag{1.3}$$

These sets stratify the coarse Kloosterman set as follows,

$$\Omega_{w_0}(c_1, c_2) = \bigsqcup_{f | \gcd(c_1, c_2)} \Omega\left(\frac{c_1}{f}, \frac{c_2}{f}, f \right). \tag{1.4}$$

The sets on the right hand side of this finer decomposition are invariant under the action of $\mathrm{U}(\mathbb{Z})$ from both sides, thus the decomposition carries over to $\mathrm{U}(\mathbb{Z})$ double-cosets. This stratification gives a decomposition of the long word SL_3 Kloosterman sum into what we call *fine Kloosterman sums*. In order to distinguish it, we denote by script $\mathcal{S}_w$:

$$\mathcal{S}_{w_0}(\mathbf{m}, \mathbf{n}; d_1, d_2, f) := \sum_{\substack{A \in \Gamma_\infty \backslash \Omega(d_1, d_2, f)/\Gamma_\infty \\ A \in u_L w_0 t(d_1 f, d_2 f) u_R}} \psi_{\mathbf{m}}(u_L) \psi_{\mathbf{n}}(u_R). \tag{1.5}$$

This finer decomposition is inspired by a reduced word decomposition of w_0 and the subsequent Bott-Samelson factorization of flag varieties. Thus we are able to write the usual (*coarse*) Kloosterman sum as a sum of *fine* Kloosterman sums,

$$S_{w_0}(\mathbf{m}, \mathbf{n}; (c_1, c_2)) = \sum_{f | \gcd(c_1, c_2)} \mathcal{S}_{w_0}\left(\mathbf{m}, \mathbf{n}; \frac{c_1}{f}, \frac{c_2}{f}, f \right). \tag{1.6}$$

1.2. **Statement of Results.** We parametrize $\Omega(d_1, d_2, f)$, thus obtaining nice expressions for $\mathcal{S}_{w_0}(\mathbf{m}, \mathbf{n}; d_1, d_2, f)$.

Theorem 1.1. *Let $n_1, n_2, m_1, m_2 \in \mathbb{Z}$. The Kloosterman sum $\mathcal{S}_{w_0}(\mathbf{m}, \mathbf{n}; d_1, d_2, f)$ is zero unless $(m_2 d_2, f) = (n_2 d_1, f)$. If this is satisfied, then the Kloosterman sum equals,*

$$f \sum_{\substack{x_3, y_3 \pmod{f} \\ x_3 y_3 \equiv 1 \pmod{f} \\ m_2 d_2 + n_2 d_1 y_3 \equiv 0 \pmod{f}}} S(n_1, (m_2 d_2 + n_2 d_1 y_3)/f; d_1) S(m_1, (n_2 d_1 + m_2 d_2 x_3)/f; d_2).$$

$$\mathcal{S}_{w_0}(\mathbf{m}, \mathbf{n}; d_1, d_2, f) = f \sum_{\substack{x, y \pmod{f} \\ xy \equiv 1 \pmod{f} \\ m_2 d_2 + n_2 d_1 y \equiv 0 \pmod{f}}} S(n_1, N(y); d_1) S(m_1, M(x); d_2). \tag{1.7}$$

Notice that when $f = 1$ this simplifies to $S(n_1, m_2 d_2; d_1) S(m_1, n_2 d_1; d_2)$.

Note that (1.6) and (1.7) together give us that $S_{w_0}(\mathbf{m}, \mathbf{n}, (c_1, c_2))$ lies in a real algebraic number field. In fact it lies in a compositum of fields of the form $\mathbb{Q}(\cos\left(\frac{2\pi}{p^k}\right))$, for various primes p and integers k.

Another application is the following explicit formula for the triple divisor sum. Let us define

$$\sigma_{s_1,s_2}(n_1,n_2) = \sum_{e_1|n_1}\sum_{e_2|n_2}\sum_{e_3|\frac{n_1e_2}{e_1}} e_1^{s_1+s_2}e_2^{s_2}e_3^{s_1}.$$

These arithmetic functions are multiplicative and show up in the Fourier coefficients of SL_3 Eisenstein series. Their values at prime powers are also related to Schur polynomials.

Substituting $n_1 = 1$ the above lemma simplifies as follows

$$\sigma_{s_1,s_2}(1,n) = \sum_{a|n} a^{s_2}\sum_{b|a} b^{s_1} = \sum_{n=e_1e_2e_3} e_1^{s_1+s_2}e_2^{s_2},$$

and in particular $d_3(n) = \sigma_{0,0}(1, n)$.

Now using the expression for the Kloosterman sum in Theorem 1.1, we write the Ramanujan sum. Compare with [Bum84, (6.3)].

Lemma 1.2. *Given $c_1, c_2 \in \mathbb{Z}^{>0}$, let us call $R_{c_1,c_2}(n_1, n_2) = S_{w_0}(\mathbf{0}, \mathbf{n}; (c_1, c_2))$ the Ramanujan sum. Then,*

$$R_{c_1,c_2}(n_1,n_2) = \sum_{\substack{f|\gcd(c_1,c_2)\\ f|\frac{n_2c_1}{f}}} f c_{c_1/f}(n_1)c_f(n_2)c_{c_2/f}\left(\frac{c_1n_2}{f^2}\right).$$

Now using the same identity as Bump [Bum84] we start to calculate the sum

$$\zeta(s_1)\zeta(s_2)\zeta(s_1+s_2-1)\sum_{c_1,c_2>0}\frac{R_{c_1,c_2}(n_1,n_2)}{c_1^{s_1}c_2^{s_2}},$$

in order to obtain $\sigma_{s_1,s_2}(n_1, n_2)$. Such equality can be justified via a study of Fourier coefficients SL_3 Eisenstein series. Yet, this is an elementary statement expressing a divisor function as a double Dirichlet series of finite exponential sums. Discovering the form of the formula took us through SL_3; however, as we see in the proof of the next proposition, an elementary proof can also be given.

Proposition 1.3. *For* $\mathrm{Re}(s_1), \mathrm{Re}(s_2) > 1$*, we have the identity*

$$\sigma_{1-s_1,1-s_2}(1,n) = \zeta(s_1)\zeta(s_2)\zeta(s_1+s_2-1)\sum_{d_1,d_2=1}^{\infty}\frac{\mu(d_1)}{d_1^{s_1}d_2^{s_2}}\sum_{f|d_1n}\frac{c_f(n)c_{d_2}(\frac{nd_1}{f})}{f^{s_1+s_2-1}}. \tag{1.8}$$

where $c_q(n)$ is the classical Ramanujan sum.

This proposition will be proved in Section 3.

Stevens, in [Ste87], has bounded the coarse long word Kloosterman sums as

$$|S_{w_0}(\mathbf{m},\mathbf{n};(c_1,c_2))| \le \tau(c_1)\tau(c_2)(m_1n_2,C)^{\frac{1}{2}}(m_2n_1,C)^{\frac{1}{2}}(c_1,c_2)^{\frac{1}{2}}\sqrt{c_1c_2}, \tag{1.9}$$

where $C = \mathrm{lcm}(c_1, c_2)$. See [But13, Theorem 4] for the above formulation.

Using the Weyl bound on the classical Kloosterman sum for the Kloosterman sum decomposition we get the following theorem.

Proposition 1.4. *Given* $\mathbf{m}, \mathbf{n} \in \mathbb{Z}^2 - (0,0)$, *and* $c_1, c_2 > 0$, *we may bound the long word* coarse *Kloosterman sum as*

$$|S_{w_0}(\mathbf{m}, \mathbf{n}; \mathbf{c})| \leq \sqrt{c_1 c_2}(c_1, c_2)^{\frac{1}{2}} \tau((c_1, c_2)) \tau(c_1) \tau(c_2) \min\{A, B\},$$

where $\tau(c)$ *is the number of divisors of* c *and*

$$A = (m_2 n_1, c_1)^{\frac{1}{2}} (n_2 m_1, c_2)^{\frac{1}{2}},$$
$$B = (m_2 n_1, c_2)^{\frac{1}{2}} (n_2 m_1, c_1)^{\frac{1}{2}}.$$

Notice that this is still stronger than (1.9) in its $\mathbf{m}$ and $\mathbf{n}$ dependence and only weaker in its $\mathbf{c}$ dependence by a very small factor of $\tau((c_1, c_2))$. This is despite the fact that in the above proof we used many potentially not sharp inequalities.

As an example we may see that using this proposition we obtain the bound

$$S_{w_0}((1,p),(1,p);(p^2,p)) = O\big(p^{5/2}\big),$$

which is sharp. The bound (1.9), on the other hand, would imply an upper bound on the order of $O_\epsilon(p^{3+\varepsilon})$.

Let $\Gamma_0(N) \subseteq \mathrm{SL}_3(\mathbb{Z})$ be the congruence subgroup consisting of matrices such that the bottom row is congruent to $\begin{pmatrix} 0 & 0 & * \end{pmatrix}$ modulo N. We note that the pieces of our stratification (1.6) encodes the level structure in a simple manner. The fine Kloosterman sums appearing in Bruggeman-Kuznetsov trace formula for the congruence group $\Gamma_0(N)$ are exactly those fine Kloosterman sums $\mathcal{S}_{w_0}(\mathbf{n}, \mathbf{m}; d_1, d_2, f)$ with $N|f$. This is a simple condition, which implies $N|c_1$ and $N|c_2$ in the notation of (1.6), but is not conversely implied by it.

1.3. **The historical background and the previous literature.** The exponential sum

$$S(m,n;c) := \sum_{\substack{a,d \pmod{c} \\ ad \equiv 1 \pmod{c}}} e\left(\frac{ma}{c} + \frac{nd}{c}\right),$$

is called the classical Kloosterman sum, first introduced by H. D. Kloosterman in [Klo27] in the context of bounding the error term arising from the circle method of G.H. Hardy, J. E. Littlewood and S. Ramanujan [HL19, HR18]. Here we use the notation $e(z) = e^{2\pi i z}$, for $z \in \mathbb{C}$.

A second context in which the Kloosterman sums appear is in exponential sums over $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$, for example in the computation of the Fourier coefficients of the classical Poincaré series.

In this second context the presence of Kloosterman sums on the *geometric side* of the Petersson and Bruggeman-Kuznetsov trace formulas, forges a connection with the spectral theory of automorphic forms. Thus Kloosterman sums are abound in works estimating L-function moments, obtaining hyperbolic equidistribution results, quantum ergodicity on hyperbolic spaces. For SL_2 automorphic forms, the Bruggeman-Kuznetsov/Petersson trace formulas have been the workhorse of virtually any result in analytic number theory concerning a family of automorphic forms and L-functions.

Given the central importance of Kloosterman sums in the rank 1 theory, attention has turned also to higher rank calculations. In the seminal work of [BFG88], the authors used

Plücker coordinates to parametrize the double cosets of the Bruhat cells of SL_3. This formulation has recently has been used in myriad applications, especially in the context of SL_3 Kuznetsov trace formula, see [Blo13], [GK13], [You16], [BBM17], [BB]. For the general higher rank case, the explicit calculation of certain Kloosterman sums in SL_r have been performed in [Fri87], [Ste87].

The work of [Fri87] notices the general rank r hyperkloosterman sum as the Kloosterman sum associated to the cyclic element $(12\cdots r)$ of the Weyl group Sym_r of SL_r. Our work shares the use of the exterior algebra in determining the coordinates of various factorizations.

1.4. **Method of Proof.** Our calculation is heavily influenced by, but does not directly use, the Bott-Samelson decomposition of a flag variety. We saw this approach first in the work of Brubaker and Friedberg in [BF15], in the context of calculating the Fourier coefficients of metaplectic Eisenstein series. Especially for the GL_3 case, the Bott-Samelson factorization has also been studied by Bump and Choie [BC14]. They have done this in the context of Schubert Eisenstein series, a new object introduced by the authors where the summation of the Eisenstein series is not over the full flag variety but over a Schubert cell. Given a Weyl group element w and $w = s_{\alpha_1}\cdots s_{\alpha_\ell}$ a reduced word decomposition of w, we can write

$$BwB = (Bs_{\alpha_1}B)(Bs_{\alpha_2}B)\cdots(Bs_{\alpha_\ell}B). \tag{1.10}$$

In fact we can accomplish this in quite a generality, see [Gar05]. Our approach in this work is to find the necessary conditions such that given an $A \in BwB \cap \mathrm{SL}_r(\mathbb{Z})$, we can write

$$\iota_{\alpha_1}(\gamma_1)\cdots\iota_{\alpha_\ell}(\gamma_\ell) \in \Gamma_\infty A \Gamma_\infty,$$

where $\gamma_i \in \mathrm{SL}_2$, in the big cell, i.e. with a nonzero lower left entry. It would be simplest if we could independently choose each $\gamma_i \in \mathrm{U}_2(\mathbb{Z})\backslash B\left(\begin{smallmatrix} & -1 \\ 1 & \end{smallmatrix}\right)B \cap \mathrm{SL}_2(\mathbb{Z})/\,\mathrm{U}_2(\mathbb{Z})$. However, the reality is subtler. In this paper, we work out the various integrality conditions and the interdependencies among the γ_i's.

1.5. **Discussion.** Historically Kloosterman introduced his sum [Klo27], in the context of the circle method applied to the sum of four squares. The problem had no Bruhat decomposition in sight. An understandable formula for a SL_3 (or higher rank) Kloosterman sum may allow researchers to recognize Kloosterman sums when they see them in their research. Thus, for the researchers working on more complicated problems involving the circle method, the exponential sums they obtain may signal to them that there may be a hidden connection to higher rank automorphic forms.

We expect that our detailed investigation into the structure of the higher rank Kloosterman sums will also lead to a refined understanding of higher rank automorphic forms. As an example, recently there has been a flurry of activity in spectral reciprocity formulas, see [BLM19], [BK19], [AK18], [Zac19], [Pet15] and of course the seminal work of Motohashi [Mot93]. These are formulas where both sides contain a moment, or a twisted moment of a family of L-functions with possibly some correction terms. One way to obtain these results is to pass from either side, perhaps via a trace formula, to a sum of exponential sums and connect these exponential sums. At this step precise and practical knowledge of the exponential sums is necessary. Great insight is to be gained from finding connections between various moments.

In a more straightforward way we also expect our results to be useful in the spectral theory of higher rank automorphic forms. Even though there have been deep results concerning higher rank automorphic forms, see [Li11], [BLM19], [LY12], these have all used the SL_2 spectral theory and Bruggeman Kuznetsov formula. The notable exceptions to these are [Blo13], [BBM17], and [You16] where the sums are over SL_3 automorphic forms. We should note however that most of these results have used only upper bounds on Kloosterman sums, and not their explicit form.

Also we can use the methods of this paper to consider the metaplectic case. As noted in [BF15] and [BBF11] the decomposition of $A = \prod_{i=1}^{r} \iota_{\alpha_i}(\left(\begin{smallmatrix} a_i & b_i \\ c_i & d_i \end{smallmatrix}\right))$ helps us easily write the Kubota symbol $\kappa(A)$ using n^{th} power residue symbols $\left(\frac{d_i}{c_i}\right)_n$ multiplicatively.

In [Mot97, Chapter 5.4, p.215] Motohashi has noted that just as the Ramanujan formula for the divisor function was used in an essential manner in obtaining the spectral formula for the fourth moment of the Riemann zeta function in [Mot93], its generalization for the triple divisor function forms a connection between the sixth moment of the Riemann zeta function and the $\mathrm{SL}_3(\mathbb{Z})$ theory, and continues to emphasize that *"...it is highly desirable to have an honest extension to* $\mathrm{SL}(3,\mathbb{Z})$ *of the theory developed in Chapters 1-3"*. Bump in [Bum84] has found such a formula, as Motohashi notes, even though this establishes the connection to the $\mathrm{SL}_3(\mathbb{Z})$ theory, the exact form of the divisor formula was not amenable to concrete calculations.

Notice that for $s_1 = s_2 = 1$ the left hand side of (1.8) is the triple divisor function $\tau_3(n) = \sum_{n_1 n_2 n_3 = n} 1$. Our formula gives a way to *expand* $\tau_3(n)$ into a double Dirichlet series of exponential sums, which hopefully can be useful in separating additive terms that appear in shifted convolution sums such as $\sum_{n \ll X} \tau_3(n)\tau_3(n+h)$.

2. Main Calculations

First, some notation.

Let V be an r dimensional vector space, with $\mathbf{e}_1, \ldots, \mathbf{e}_r$ as standard basis vectors. Given an element $A \in \mathrm{GL}_r$ the action of A on elements of the k-fold wedge product are defined as

$$A(v_1 \wedge v_2 \wedge \cdots \wedge v_k) = (Av_1) \wedge (Av_2) \wedge \cdots \wedge (Av_k).$$

For subsets $I = \{i_1 < i_2 < \ldots < i_k\} \subseteq \{1, \ldots, r\}$ the vectors $\mathbf{e}_I := \mathbf{e}_{i_1} \wedge \cdots \wedge \mathbf{e}_{i_k}$, form a basis of $\bigwedge^k V$. The action of A is calculated explicitly via the minors as,

$$A\mathbf{e}_I = \sum_{\substack{J \subseteq \{1,\ldots,r\} \\ |J|=k}} M_{I,J} \mathbf{e}_J.$$

Writing $\mathbf{e}_J^* := \mathbf{e}_{j_1}^* \wedge \mathbf{e}_{j_2}^* \wedge \cdots \wedge \mathbf{e}_{j_k}^* \in (\bigwedge^k V)^* \cong \bigwedge^k V^*$, where $\mathbf{e}_1^*, \ldots, \mathbf{e}_r^*$ are the dual standard basis elements of V^*, we can also write $M_{I,J} = \langle \mathbf{e}_J^*, A\mathbf{e}_I \rangle$.

2.1. **Reduced Word Decomposition and the parametrization of the fine Kloosterman cells.** In the symmetric group S_3, let us call the simple transpositions $s_\alpha = (12), s_\beta = (23)$. Using the reduced word decomposition $w_0 = s_\alpha s_\beta s_\alpha$ we parametrize the fine Kloosterman sets, that is, given

$$\gamma_2 = \begin{pmatrix} x_2 & b_2 \\ d_2 & y_2 \end{pmatrix}, \quad \gamma_3 = \begin{pmatrix} x_3 & D \\ f & y_3 \end{pmatrix}, \quad \gamma_1 = \begin{pmatrix} x_1 & b_1 \\ d_1 & y_1 \end{pmatrix}, \tag{2.1}$$

we use the product

$$\iota_\alpha(\gamma_2)\iota_\beta(\gamma_3)\iota_\alpha(\gamma_1) \tag{2.2}$$

to express elements of $\Omega(d_1, d_2, f)$.

Every product of the form (2.2) with the matrices (2.1) in $\mathrm{SL}_2(\mathbb{Z})$ gives an element of $\Omega(d_1, d_2, f)$. It is, however not true that any element of $\Omega(d_1, d_2, f)$ can be expressed as such a product. Firstly it is sometimes necessary to pick the matrices (2.1) in $\mathrm{SL}_2(\mathbb{Q})$, and secondly some matrices A cannot be obtained in such a manner as can see by taking a matrix with $D = 0$ in the notation of Proposition 2.1 below. However it is possible to find a representative $A' \in \Gamma_\infty A \Gamma_\infty$ that factorizes.

Call $A_{33}M_{33} - 1 = fD$. D is an integer. Multiplying with an element of $\mathrm{U}_3(\mathbb{Z})$ element on either side we can make sure that $D \neq 0$.

Given a vector space V with a three dimensional basis, and using the action of $A = u_L w_0 t u_R$, on various basis elements $\mathbf{e}_I$ of the exterior algebra $\bigwedge V$, one obtains this explicit Bruhat decomposition

$$A = \begin{pmatrix} 1 & M_{23}/M_{13} & A_{11}/A_{31} \\ & 1 & A_{21}/A_{31} \\ & & 1 \end{pmatrix} w_0 \begin{pmatrix} A_{31} & & \\ & \frac{M_{13}}{A_{31}} & \\ & & \frac{1}{M_{13}} \end{pmatrix} \begin{pmatrix} 1 & A_{32}/A_{31} & A_{33}/A_{31} \\ & 1 & M_{12}/M_{13} \\ & & 1 \end{pmatrix}. \tag{2.3}$$

comparing coordinates from both sides of the action.

Proposition 2.1. *Let A be an integral matrix in the big Bruhat cell. Assume (by changing to a different element in the double coset $\mathrm{U}_3(\mathbb{Z})A\,\mathrm{U}_3(\mathbb{Z})$ if necessary) that $fD := A_{33}M_{33} - 1 \neq 0$. We have the explicit decomposition,*

$$\begin{aligned} A = e_\alpha\left(\frac{M_{23}/f}{d_2}\right) s_\alpha h_\alpha(d_2) e_\alpha\left(\frac{A_{23}/D}{d_2}\right) e_\beta\left(\frac{M_{33}}{f}\right) s_\beta h_\beta(f) \\ \times e_\beta\left(\frac{A_{33}}{f}\right) e_\alpha\left(\frac{M_{32}/D}{d_1}\right) s_\alpha h_\alpha(d_1) e_\alpha\left(\frac{A_{32}/f}{d_1}\right). \end{aligned}$$

This proposition states that the double cosets

$$\mathrm{U}_3(\mathbb{Z})\iota_\alpha(\gamma_2)\iota_\beta(\gamma_3)\iota_\alpha(\gamma_1)\,\mathrm{U}_3(\mathbb{Z}),$$

with γ_i as in (2.1) with $b_i = \frac{x_i y_i - 1}{d_i}$ for $i = 1, 2$ and $b_3 = \frac{x_3 y_3 - 1}{f} = D$ and $x_2, y_1, x_3, y_3, b_3 \in \mathbb{Z}$ and $x_1, y_2 \in \frac{1}{D}\mathbb{Z}$ gives a surjective map onto $\Gamma_\infty \backslash \Omega(d_1, d_2, f)/\Gamma_\infty$. Furthermore it is enough to take a single representative $y_1 \pmod{d_1}, x_2 \pmod{d_2}$ and $x_3, y_3 \pmod{f}$.

We omit the details of the proof. The result is achieved by first assuming that A is of the form $\iota_\alpha(\gamma_2)\iota_\beta(\gamma_3)\iota_\alpha(\gamma_1)$ with the coordinates of γ_2, γ_3 and γ_1 as in (2.1). Using the word-based factorization coordinates we calculate the action on various basis elements of the exterior algebra. We get that $A_{31} = d_1 f$, $M_{13} = d_2 f$, as well as,

$$
\begin{aligned}
x_2 &= \frac{\langle \mathbf{e}_{1,3}^*, A\mathbf{e}_{1,2}\rangle}{f} = \frac{\begin{vmatrix} A_{11} & A_{12} \\ A_{31} & A_{32} \end{vmatrix}}{f}, & y_2 &= \frac{\langle \mathbf{e}_2^*, A\mathbf{e}_3\rangle}{D} = \frac{A_{23}}{D}, \\
x_3 &= \langle \mathbf{e}_{1,2}^*, A\mathbf{e}_{1,2}\rangle = \begin{vmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{vmatrix}, & y_3 &= \langle \mathbf{e}_3^*, A\mathbf{e}_3\rangle = A_{33}, \\
x_1 &= \frac{\langle \mathbf{e}_{1,2}^*, A\mathbf{e}_{1,3}\rangle}{D} = \frac{\begin{vmatrix} A_{11} & A_{13} \\ A_{21} & A_{23} \end{vmatrix}}{D}, & y_1 &= \frac{\langle \mathbf{e}_3^*, A\mathbf{e}_2\rangle}{f} = \frac{A_{32}}{f}.
\end{aligned} \tag{2.4}
$$

Also from the fact that f divides A_{32} and M_{23} we deduce that $x_2, x_3, y_3, y_1 \in \mathbb{Z}$ and $y_2, x_1 \in \frac{1}{D}\mathbb{Z}$.

Multiplying these gets A back, justifying our assumption.

Let us now express the coordinates of the Bruhat decomposition using these coordinates. So we write $A = u_L w_0 t u_R$ and also $A = \iota_\alpha(\gamma_2)\iota_\beta(\gamma_3)\iota_\alpha(\gamma_1)$.

From (2.3) we know that

$$
u_L = \begin{pmatrix} 1 & \frac{\langle \mathbf{e}_{1,3}^*, A\mathbf{e}_{1,2}\rangle}{t_1 t_2} & \frac{\langle \mathbf{e}_1^*, A\mathbf{e}_1\rangle}{t_1} \\ & 1 & \frac{\langle \mathbf{e}_2^*, A\mathbf{e}_1\rangle}{t_1} \\ & & 1 \end{pmatrix} \quad \text{and} \quad u_R = \begin{pmatrix} 1 & \frac{\langle \mathbf{e}_3^*, A\mathbf{e}_2\rangle}{t_1} & \frac{\langle \mathbf{e}_3^*, A\mathbf{e}_3\rangle}{t_1} \\ & 1 & \frac{\langle \mathbf{e}_{2,3}^*, A\mathbf{e}_{1,3}\rangle}{t_1 t_2} \\ & & 1 \end{pmatrix}.
$$

Denoting $u = x_1 d_2 + y_2 x_3 d_1$, and $v = x_1 y_3 d_2 + y_2 d_1$, we have

$$
\begin{aligned}
\langle \mathbf{e}_{1,3}^*, A\mathbf{e}_{1,2}\rangle &= x_2 f, & \langle \mathbf{e}_{2,3}^*, A\mathbf{e}_{1,3}\rangle &= x_1 y_3 d_2 + d_1 y_2 = v, \\
\langle \mathbf{e}_3^*, A\mathbf{e}_2\rangle &= y_1 f, & \langle \mathbf{e}_2^*, A\mathbf{e}_1\rangle &= x_1 d_2 + y_2 x_3 d_1 = u,
\end{aligned}
$$

and

$$
\langle \mathbf{e}_1^*, A\mathbf{e}_1\rangle = (x_1 x_2 + x_3 b_2 d_1) = \frac{x_2 u - x_3 d_1}{d_2}, \qquad \langle \mathbf{e}_3^*, A\mathbf{e}_3\rangle = y_3.
$$

Notice that $u, v \in \mathbb{Z}$. Combining these calculations, we obtain the following result.

Proposition 2.2. *Given a matrix $A \in \mathrm{SL}_3(\mathbb{Z})$, choose d_1, d_2, f as in* (1.3)*. After replacing A with $A' \sim A$ if necessary, we can write $A = \iota_\alpha(\gamma_2)\iota_\beta(\gamma_3)\iota_\alpha(\gamma_1)$ with γ_i as in* (2.1)*, and $u, v \in \mathbb{Z}$ as above its Bruhat decomposition has the coordinates*

$$
A = \begin{pmatrix} 1 & \frac{x_2}{d_2} & \frac{x_2 u - x_3 d_1}{d_1 d_2 f} \\ & 1 & \frac{u}{d_1 f} \\ & & 1 \end{pmatrix} \begin{pmatrix} & & 1 \\ & -1 & \\ 1 & & \end{pmatrix} \begin{pmatrix} d_1 f & & \\ & \frac{d_2}{d_1} & \\ & & \frac{1}{d_2 f} \end{pmatrix} \begin{pmatrix} 1 & \frac{y_1}{d_1} & \frac{y_3}{d_1 f} \\ & 1 & \frac{v}{d_2 f} \\ & & 1 \end{pmatrix}, \tag{2.5}
$$

with all the visible parameters integral, x_2, y_1, x_3, y_3 relatively prime to $d_1 d_2 f$, and $x_3 y_3 \equiv 1 \pmod{f}$.

In the next proposition we give the conditions under which the coordinates in (2.5) give rise to the same $\mathrm{U}_3(\mathbb{Z})$–double coset.

Proposition 2.3. *Given nonzero integers d_1, d_2, f and $y_1 \in (\mathbb{Z}/d_1\mathbb{Z})^*$, $x_2 \in (\mathbb{Z}/d_2\mathbb{Z})^*$ and $x_3, y_3 \in \mathbb{Z}/f\mathbb{Z}$ satisfying $x_3 y_3 \equiv 1 \pmod{f}$, the product in* (2.5) *gives rise to an integral* *matrix if and only if the following congruence conditions are satisfied:*

$$ux_2 \equiv d_1x_3 \pmod{d_2}, \tag{2.6}$$

$$uy_1 \equiv d_2 \pmod{d_1}, \tag{2.7}$$

$$ux_2y_1 \equiv d_1x_3y_1 + d_2x_2 \pmod{d_1d_2}, \tag{2.8}$$

$$v \equiv uy_3 \pmod{d_1f}, \tag{2.9}$$

$$vx_2 \equiv uy_3x_2 + d_1(1-x_3y_3) \pmod{d_1d_2f}. \tag{2.10}$$

Furthermore a matrix B that formed in the same way from the coordinates Y_1, X_2, U, V and x_3, y_3 is in $\mathrm{U}_3(\mathbb{Z})A\,\mathrm{U}_3(\mathbb{Z})$ if and only if

$$\begin{aligned} y_1 &\equiv Y_1 \pmod{d_1},\\ x_2 &\equiv X_2 \pmod{d_2},\\ u &\equiv U \pmod{d_1d_2f},\\ v &\equiv V \pmod{d_1d_2f}. \end{aligned}$$

Remark 1. *If we choose y_1, x_2 to be relatively prime to d_1d_2f (which we can via switching to a different matrix in the $\mathrm{U}(\mathbb{Z})$ double coset if necessary) then* (2.8) *and* (2.10) *imply the remaining congruence relations.*

From now on we will assume x_2 and y_1 are chosen to be relatively prime to d_1d_2f.

Since the equation (2.8) determines u up to d_1d_2 but u determines the double coset up to modulo d_1d_2f, the set of allowed solutions are

$$u \equiv d_1x_3\overline{x_2} + d_2\overline{y_1} + d_1d_2k \pmod{d_1d_2f}, \tag{2.11}$$

with $k \in \mathbb{Z}/f\mathbb{Z}$.

This then determines $v \pmod{d_1d_2f}$ completely and we have for each such u,

$$v \equiv (d_1x_3\overline{x_2} + d_2\overline{y_1} + d_1d_2k)y_3 + d_1(1-x_3y_3)\overline{x_2} \equiv d_2\overline{y_1}y_3 + d_1\overline{x_2} + d_1d_2y_3k \pmod{d_1d_2f}.$$

This gives a parametrization of the fine Kloosterman cells.

Corollary 2.4. *Let d_1, d_2, f be nonzero integers, and fix the sets $\mathcal{Y}_{d_1}$ and $\mathcal{X}_{d_1}$, a complete set of reduced residue class representatives $y_1 \pmod{d_1}^*, x_2 \pmod{d_2}^*$ such that x_2, y_1, are relatively prime to d_1d_2f. Let $\mathcal{F}_f = \{(x_3,y_3) \in \{f+1,\dots,2f\} | x_3y_3 \equiv 1 \pmod{f}\}$ and let $k \in \mathcal{K}_f$ simply run through integers from 0 to $f-1$. There is a bijection*

$$\mathcal{X}_{d_2}\times\mathcal{Y}_{d_1}\times\mathcal{F}_f\times\mathcal{K}_F \longrightarrow \mathrm{U}_3(\mathbb{Z})\backslash\Omega(d_1,d_2,f)/\,\mathrm{U}_3(\mathbb{Z})$$

$$(x_2,y_1,(x_3,y_3),k) \longmapsto \mathrm{U}_3(\mathbb{Z})\begin{pmatrix} \frac{ux_2-d_1x_3}{d_2} & \frac{ux_2y_1-d_1x_3y_1-x_2d_2}{d_1d_2} & \frac{-vx_2+ux_2y_3+d_1(1-x_3y_3)}{d_1d_2f}\\ u & \frac{uy_1-d_2}{d_1} & \frac{uy_3-v}{d_1f}\\ d_1f & fy_1 & y_3 \end{pmatrix}\mathrm{U}_3(\mathbb{Z}),$$

where $u = d_1x_3\overline{x_2} + d_2\overline{y_1} + d_1d_2k$ and $v = d_2\overline{y_1}y_3 + d_1\overline{x_2} + d_1d_2y_3k$.

Remark 2. *The condition that $f < x_3, y_3 < 2f$ is not important. Any fixed set of reduced residue classes would work as long as $x_3y_3 - 1 \neq 0$.*

Corollary 2.5. *The number of elements in* $\Omega_{w_0}(c_1, c_2)$ *is given by*

$$|\Gamma_\infty\backslash\Omega_{w_0}(c_1,c_2)/\Gamma_\infty| = \sum_{f|(c_1,c_2)} \phi\left(\frac{c_1}{f}\right)\phi\left(\frac{c_2}{f}\right)\phi(f)f.$$

2.2. **Evaluation of Fine Kloosterman Sums.** According to this parametrization we evaluate $\mathcal{S}_{w_0}(\mathbf{m},\mathbf{n};d_1,d_2,f)$. The k sum will give us a restriction on the set of (x_3,y_3) pairs as well as the condition that $(n_2d_1,f)=(m_2d_2,f)$.

Proof of Theorem 1.1. We calculate by using the definition of the fine Kloosterman sum, the coordinatization of the Kloosterman set from 2.4, and the explicit form of the superdiagonal elements in the unipotent factors of the Bruhat decomposition in terms of these coordinates as in (2.5),

$$\begin{aligned}\mathcal{S}_{w_0}(\mathbf{m},\mathbf{n};d_1,d_2,f) &= \sum_{\substack{\gamma\in\Gamma_\infty\backslash\Omega(d_1,d_2,f)/\Gamma_\infty\\ \gamma\in u_Lw_0t(d_1f,d_2f)u_R}} \psi_{(m_1,m_2)}(u_L)\psi_{(n_1,n_2)}(u_R)\\ &= \sum_{\substack{x_2\in\mathcal{X}_{d_2}\\ y_1\in\mathcal{Y}_{d_1}}}\sum_{(x_3,y_3)\in\mathcal{F}_f}\sum_{k=0}^{f-1} e\left(\frac{m_1x_2}{d_2}+\frac{m_2u}{d_1f}+\frac{n_1y_1}{d_1}+\frac{n_2v}{d_2f}\right).\end{aligned}$$

Then we plug in the values for u and v in terms of the given coordinates,

$$\begin{aligned}\mathcal{S}_{w_0}(\mathbf{m},\mathbf{n};d_1,d_2,f) = &\sum_{\substack{x_2\in\mathcal{X}_{d_2}\\ y_1\in\mathcal{Y}_{d_1}}}\sum_{(x_3,y_3)\in\mathcal{F}_f} e\left(\frac{m_1x_2}{d_2}+\frac{n_2d_1\overline{x_2}}{d_2f}+\frac{m_2x_3\overline{x_2}}{f}\right)\\ &\times e\left(\frac{n_1y_1}{d_1}+\frac{m_2d_2\overline{y_1}}{d_1f}+\frac{n_2y_3\overline{y_1}}{f}\right)\sum_{k=0}^{f-1} e\left(\frac{m_2d_2+n_2d_1y_3}{f}k\right).\end{aligned}$$

The innermost sum over k gives us the congruence condition

$$m_2d_2+n_2d_1y_3\equiv 0 \pmod f, \tag{2.12}$$

for otherwise the sum vanishes. Some $y_3\in(\mathbb{Z}/f\mathbb{Z})^*$ satisfies this if and only if $(m_2d_2,f)=(n_2d_1,f)$. Thus,

$$\begin{aligned}\mathcal{S}_{w_0}(\mathbf{m},\mathbf{n};d_1,d_2,f) = f &\sum_{\substack{(x_3,y_3)\in\mathcal{F}_f\\ m_2d_2+n_2d_1y_3\equiv 0 \pmod f}}\sum_{y_1\in\mathcal{Y}_{d_1}} e\left(\frac{n_1y_1}{d_1}+\frac{(m_2d_2+n_2d_1y_3)\overline{y_1}}{d_1f}\right)\\ &\times\sum_{x_2\in\mathcal{X}_{d_2}} e\left(\frac{m_1x_2}{d_2}+\frac{(m_2d_2x_3+n_2d_1)\overline{x_2}}{d_2f}\right).\end{aligned}$$

Let y_3 be chosen so that (2.12) is satisfied. Define the integers $N=N(y_3):=(m_2d_2+n_2d_1y_3)/f$ and $M=M(x_3):=(m_2d_2x_3+n_2d_1)/f$. These are both integers, due to the condition on (x_3,y_3). Note that if $x_3\equiv x_3' \pmod f$ then $M(x_3)\equiv M(x_3') \pmod{d_1}$ and similarly for $N(y_3)$. The fine Kloosterman sum is

$$\mathcal{S}_{w_0}(\mathbf{m},\mathbf{n};d_1,d_2,f) = f \sum_{\substack{x_3,y_3 \pmod f \\ x_3y_3\equiv 1 \pmod f \\ m_2d_2+n_2d_1y_3\equiv 0 \pmod f}} S(n_1,N(y_3);d_1)S(m_1,M(x_3);d_2). \qquad \square$$

Let us show just how explicitly we can calculate *coarse* Kloosterman sums using the above result. We calculate, for an odd prime p,

$$S_{w_0}((1,p),(1,p);(p^2,p)) = \mathcal{S}_{w_0}((1,p),(1,p);p^2,p,1) + \mathcal{S}_{w_0}((1,p),(1,p);p,1,p).$$

The first term with $f=1$ is easy to calculate, we can take $x_3=y_3=0$ in (1.7) and get,

$$\mathcal{S}_{w_0}((1,p),(1,p);p^2,p,1) = S(1,p^2;p)S(1,p^3;p) = \mu(p^2)\mu(p) = 0.$$

The second fine Kloosterman sum can be evaluated as

$$\mathcal{S}_{w_0}((1,p),(1,p);p,1,p) = p \sum_{\substack{x_3y_3\equiv 1 \pmod p \\ p+y_3p^2\equiv 0 \pmod p}} S(1,\tfrac{p+p^2y_3}{p};p)S(1,\tfrac{px_3+p^2}{p};1) = p(p-1)S(1,1;p),$$

since $(p-1)$ many (x_3,y_3) pairs all yield the same answer. Thus we get

$$\mathcal{S}_{w_0}((1,p),(1,p);(p^2,p)) = p(p-1)S(1,1;p). \tag{2.13}$$

Another example would be $S_{w_0}((1,1),(p,p),(p,p)) = 2-p$.

Finally let's take integers m_1,m_2,n_1,n_2 all coprime to p.

$$S_{w_0}(\mathbf{m},\mathbf{n};(p,p)) = \mathcal{S}_{w_0}(\mathbf{m},\mathbf{n};p,p,1) + \mathcal{S}_{w_0}(\mathbf{m},\mathbf{n};1,1,p).$$

The $f=1$ case is simply $S(n_1,m_2p;p)S(m_1,n_2p;p) = c_p(n_1)c_p(m_1) = \mu(p)^2 = 1$ and the $f=p$ case is $pS(n_1,(m_2+n_2y_3)/p;1)S(m_1,(m_2x_3+n_2)/p;1)$ for the unique (x_3,y_3) pair modulo p, that makes the second arguments integers. Thus we get p. Together we get the identity [BB, (1.3)], i.e. that $S(\mathbf{m},\mathbf{n};(p,p)) = p+1$.

3. Proofs

We now include proofs of statements made in the Section 1.2

Proof of Lemma 1.2. Simply by using (1.6) and Theorem 1.1, we write,

$$\begin{aligned} R_{c_1,c_2}(n_1,n_2) &= \sum_{f\mid \gcd(c_1,c_2)} f\mathcal{S}_{w_0}(\mathbf{0},\mathbf{n};\frac{c_1}{f},\frac{c_2}{f},f) \\ &= \sum_{\substack{f\mid \gcd(c_1,c_2) \\ f\mid n_2d_1}} \sum_{\substack{x_3 \pmod f \\ \gcd(x_3,f)=1}} fS\left(n_1,\frac{n_2d_1y_3}{f},d_1\right)S\left(0,\frac{n_2d_1}{f};d_2\right). \end{aligned}$$

Here $d_1 := \frac{c_1}{f}$ and $d_2 := \frac{c_2}{f}$. We can evaluate the y_3 sum as

$$\begin{aligned}\sideset{}{^*}\sum_{y_3 \pmod f} S\left(n_1, \frac{n_2 d_1 y_3}{f}, d_1\right) &= \sideset{}{^*}\sum_{u \pmod{d_1}} e\left(\frac{n_1 u}{d_2}\right) \sideset{}{^*}\sum_{x_3 \pmod f} e\left(\frac{n_2 \overline{u} x_3}{f}\right)\\ &= \sideset{}{^*}\sum_{u \pmod{d_1}} e\left(\frac{n_1 u}{d_2}\right) c_f(m_2 \overline{u})\\ &= c_{d_2}(n_1) c_f(n_2).\end{aligned}$$

In the last line, we used the fact that $c_f(m_1\overline{u}) = c_f(m_1)$. We can do this because we have freedom to choose $\overline{u}$ as any element of the reduced residue classes $\pmod{d_2}$, so $\overline{u}$ can be a large prime, and in particular we can assume $\overline{u}$ is an integer relatively prime to f. This gives the result. □

Elementary proof of Proposition 1.3. In this proof we use the simplified notation $(a, b) = \gcd(a, b)$.

Substituting the form of the general Ramanujan sum from Lemma 1.2, we start our calculation

$$\begin{aligned}&\zeta(s_1)\zeta(s_2)\zeta(s_1+s_2-1)\sum_{d_1,d_2=1}^{\infty}\frac{1}{d_1^{s_1}d_2^{s_2}}\sum_{f|n_2d_1}\frac{c_{d_1}(n_1)c_f(n_2)c_{d_2}(n_2d_1/f)}{f^{s_1+s_2-1}}\\ &=\zeta(s_1)\zeta(s_1+s_2-1)\sum_{d_1=1}^{\infty}\frac{1}{d_1^{s_1}}\sum_{f|n_2d_1}\frac{c_{d_1}(n_1)c_f(n_2)\sigma_{1-s_2}(n_2d_1/f)}{f^{s_1+s_2-1}}.\end{aligned}$$

Here we used the classical Ramanujan identity (i.e. $\sigma_{1-s}(n) = \zeta(s)\sum_{\ell=1}^{\infty} c_\ell(n)\ell^{-s}$) on the d_2-sum. Let us assume $n_1 = 1$ now, so that $c_{d_1}(n_1) = \mu(d_1)$. Also put $(f, n_2) = e$. This gives $\gcd(n_2/e, f/e) = 1$, and so $\frac{f}{e}|d_1$. Changing variables $f/e \mapsto f$ we have,

$$\begin{aligned}&\zeta(s_1)\zeta(s_1+s_2-1)\sum_{e|n_2}\frac{1}{e^{s_1+s_2-1}}\sum_{d_1=1}^{\infty}\frac{\mu(d_1)}{d_1^{s_1}}\sum_{\substack{f|d_1\\(f,n_2/e)=1}}\frac{c_{fe}(n_2)\sigma_{1-s_2}(\frac{n_2}{e}\frac{d_1}{f})}{f^{s_1+s_2-1}}\\ &=\zeta(s_1)\zeta(s_1+s_2-1)\sum_{e|n_2}\frac{1}{e^{s_1+s_2-1}}\sum_{g|e}\mu(\tfrac{e}{g})g\sum_{d_1=1}^{\infty}\frac{\mu(d_1)}{d_1^{s_1}}\sum_{\substack{f|d_1\\(f,n_2/g)=1}}\frac{\mu(f)\sigma_{1-s_2}(\frac{n_2}{e}\frac{d_1}{f})}{f^{s_1+s_2-1}}.\end{aligned}$$

Here we inserted $c_q(n) = \sum_{g|(q,n)}\mu(\frac{q}{g})g$, noting that $(fe, n_2) = e$. The coefficient of the d_1-Dirichlet series is almost a multiplicative function. We note that for a fixed n the function $\sigma_\alpha(nd)/\sigma_\alpha(d)$ is a truly multiplicative function of d. Exchanging the order of the e and g sum we obtain,

$$\zeta(s_1)\zeta(s_1+s_2-1)\sum_{g|n_2}\frac{1}{g^{s_1+s_2-2}}\sum_{e|n_2/g}\frac{\mu(e)\sigma_{1-s_2}(\frac{n_2/g}{e})}{e^{s_1+s_2-1}}\sum_{d_1=1}^{\infty}\frac{\mu(d_1)}{d_1^{s_1}}\sum_{\substack{f|d_1\\(f,n_2/g)=1}}\frac{\mu(f)\sigma_{1-s_2}(\frac{n_2/g}{e}\frac{d_1}{f})}{f^{s_1+s_2-1}\sigma_{1-s_2}(\frac{n_2/g}{e})}.$$

For a cleaner notation we drop the subscripts at this point, and write d, n. Using the fact that the coefficients of the Dirichlet seris in the d-variable are multiplicative, this sum equals

$$\zeta(s_1)\zeta(s_1+s_2-1)\sum_{g|n}\frac{1}{g^{s_1+s_2-2}}\sum_{e|n/g}\frac{\mu(e)\sigma_{1-s_2}(\frac{n/g}{e})}{e^{s_1+s_2-1}}$$
$$\times\prod_{q\nmid\frac{n}{g}}\left(1-\frac{1}{q^{s_1}}\left(\sigma_{1-s_2}(q)-\frac{1}{q^{s_1+s_2-1}}\right)\right)\prod_{p|\frac{n}{g}}\left(1-\frac{1}{p^{s_1}}\frac{\sigma_{1-s_2}(\frac{n/g}{e}p)}{\sigma_{1-s_2}(\frac{n/g}{e})}\right).$$

The q factor is

$$1-\frac{1}{q^{s_1}}-\frac{1}{q^{s_1+s_2-1}}+\frac{1}{q^{2s_1+s_2-1}}=\left(1-\frac{1}{q^{s_1}}\right)\left(1-\frac{1}{q^{s_1+s_2-1}}\right),$$

which cancel with the Euler factors of the two zeta functions.

So let us assume $n=p^k$. If $g=n$ then the e sum is simply $1=\sigma_{1-s_2}(n/g)$. Now if $g\neq n$, we have the e sum as,

$$\left(\sigma_{1-s_2}(n/g)-\frac{\sigma_{1-s_2}(pn/g)}{p^{s_1}}-\frac{\sigma_{1-s_2}(\frac{n/g}{p})}{p^{s_1+s_2-1}}+\frac{\sigma_{1-s_2}(n/g)}{p^{2s_1+s_2-1}}\right).$$

We then apply the Hecke relation for divisor sums, i.e. that if $p|n$,

$$\sigma_\alpha(np)=\sigma_\alpha(n)\sigma_\alpha(p)-p^\alpha\sigma_\alpha(n/p).$$

Thus we have

$$\left(\sigma_{1-s_2}(n/g)-\frac{\sigma_{1-s_2}(n/g)}{p^{s_1}}\sigma_{1-s_2}(p)+\frac{\sigma_{1-s_2}(\frac{n/g}{p})}{p^{s_1+s_2-1}}-\frac{\sigma_{1-s_2}(\frac{n/g}{p})}{p^{s_1+s_2-1}}+\frac{\sigma_{1-s_2}(n/g)}{p^{2s_1+s_2-1}}\right)$$
$$=\sigma_{1-s_2}(n/g)\left(1-\frac{1}{p^{s_1}}\left(1+\frac{1}{p^{s_2-1}}\right)+\frac{1}{p^{2s_1+s_2-1}}\right)$$
$$=\zeta_p(s_1)\zeta_p(s_1+s_2-1)\sigma_{1-s_2}(n/g).$$

Here $\zeta_p(s)=(1-p^{-s})^{-1}$, is the p-Euler factor, that cancels with the Riemann zeta function.

Therefore we obtain that the whole sum is, $\sum_{g|n}\frac{1}{g^{s_1+s_2-2}}\sigma_{1-s_2}(n/g)$. □

Proof of Proposition 1.4. Given the decomposition of Kloosterman sum as a sum of product of two classical Kloosterman sums as in Theorem 1.1, and using the Weyl bound on individual terms,

$$\begin{aligned}
&|S_{w_0}(\mathbf{m},\mathbf{n};(c_1,c_2))|\\
&\leq\sum_{f|(c_1,c_2)}f\sum_{\substack{x_3y_3\equiv1\pmod f\\ m_2d_2+y_3n_2d_1\equiv0\pmod f}}(n_1,d_1)^{\frac12}(m_1,d_2)^{\frac12}\sqrt{d_1d_2}\tau(d_1)\tau(d_2)\\
&\leq\sum_{\substack{f|(c_1,c_2)\\(m_2d_2,f)=(n_2d_1,f)}}(f,m_2d_2)(n_1,d_1)^{\frac12}(m_1,d_2)^{\frac12}\sqrt{c_1c_2}\tau(d_1)\tau(d_2).
\end{aligned}\tag{3.1}$$

Here $d_i = c_i/f$ and we bounded the number of solutions to the congruence equation $m_2d_2 + y_3n_2d_1 \equiv 0 \pmod f$ with $y_3 \in (\mathbb{Z}/f\mathbb{Z})^*$ by simply (m_2d_2, f).

Notice that this answer is not symmetric in the variables. However the decomposition of S_w into the stratification induced by $w_0 = s_\beta s_\alpha s_\beta$ comes to the rescue.

Given $A = \iota_\alpha(\gamma_2)\iota_\beta(\gamma_3)\iota_\alpha(\gamma_1)$ the involution $A^\dagger := w_0({}^tA^{-1})w_0^{-1}$ is a homomorphism fixing $\mathrm{U}(\mathbb{Z})$, and therefore it preserves $\mathrm{U}(\mathbb{Z})$–double cosets. This involution does not preserve our finer decomposition, however it sends the stratification based on one reduced word decomposition to the other. Indeed $A^\dagger = \iota_\beta(\gamma_2)\iota_\alpha(\gamma_3)\iota_\beta(\gamma_1)$. The entries of $A^\dagger$ are given by $A^\dagger = \left(\begin{smallmatrix} M_{33} & M_{32} & M_{31} \\ M_{23} & M_{22} & M_{21} \\ M_{13} & M_{12} & M_{11}\end{smallmatrix}\right)$. The Kloosterman sums based on this fine decomposition are written the same way except we exchange $m_1 \leftrightarrow m_2$, $n_1 \leftrightarrow n_2$ and $d_1 \leftrightarrow d_2$. So we get

$$S_{w_0}(\mathbf{m}, \mathbf{n}; (c_1, c_2)) \leq \sum_{\substack{f|(c_1,c_2) \\ (n_1d_2,f)=(m_1d_1,f)}} (f, m_1d_1)(m_2, d_1)^{\frac{1}{2}}(n_2, d_2)^{\frac{1}{2}}\sqrt{c_1c_2}\tau(d_1)\tau(d_2). \tag{3.2}$$

Since we are adding over f such that $(m_2d_2, f) = (n_2d_1, f)$, we write in (3.1),

$$(f, m_2d_2)^2 = (f, m_2d_2)(f, n_2d_1) = (f, d_2)(f, d_1)(\tfrac{f}{(f,d_2)}, m_2)(\tfrac{f}{(f,d_1)}, n_2)).$$

Combining this with $(\frac{f}{(d_2,f)}, m_2)(d_1, n_1) \leq (d_1f, m_2n_1) = (c_1, m_2n_1)$, and similarly with $(\frac{f}{(d_1,f)}, n_2)(d_2, m_1) \leq (c_2, m_1n_2)$ we get the term $\sqrt{(d_1, f)(d_2, f)}A$. Assume that $c_1 = p^k$ and $c_2 = p^\ell$ with $\ell \leq k$. Then as f runs through powers of p, the maximum value of $(d_1, f)(d_2, f)$ is achieved for $f = p^r$ with $\frac{\ell}{2} \leq r \leq \frac{k}{2}$ and that value is $< p^\ell$. By multiplicativity we get that,

$$\max_{f|(c_1,c_2)} (\tfrac{c_1}{f}, f)(\tfrac{c_2}{f}, f) \leq (c_1, c_2).$$

There are at most $\tau((c_1, c_2))$ many summands. This gives us the bound with A. Starting with (3.2) instead, we get the bound with B. Considered together, we obtain the given statement. $\square$

Acknowledgements. We thank Prof. Masatoshi Suzuki for organizing the RIMS analytic number theory conference, giving the first author the opportunity to speak at the conference which this proceeding is based on. We would also like to thank RIKEN iTHEMS for providing the first author with office space. We thank Matthew Young, Valentin Blomer and Daniel Bump for useful comments. We would especially like to thank Solomon Friedberg and Jack Buttcane for catching a significant error in an earlier draft of [KN20].

References

[AK18] Nickolas Andersen and Eren Mehmet Kıral. Level reciprocity in the twisted second moment of Rankin-Selberg L-functions. *Mathematika*, 64(3):770–784, 2018.

[BB] Valentin Blomer and Jack Buttcane. Global decomposition of GL(3) Kloosterman sums and the spectral large sieve. *Journal für die reine und angewandte Mathematik (Crelles Journal).*

[BBF11] Ben Brubaker, Daniel Bump, and Solomon Friedberg. Weyl group multiple Dirichlet series, Eisenstein series and crystal bases. *Ann. of Math. (2)*, 173(2):1081–1120, 2011.

[BBM17] Valentin Blomer, Jack Buttcane, and Péter Maga. Applications of the Kuznetsov formula on GL(3) II: the level aspect. *Math. Ann.*, 369(1-2):723–759, 2017.

[BC14] Daniel Bump and YoungJu Choie. Schubert Eisenstein series. *Amer. J. Math.*, 136(6):1581–1608, 2014.

[BF15] Benjamin Brubaker and Solomon Friedberg. Whittaker coefficients of metaplectic Eisenstein series. *Geom. Funct. Anal.*, 25(4):1180–1239, 2015.

[BFG88] Daniel Bump, Solomon Friedberg, and Dorian Goldfeld. Poincaré series and Kloosterman sums for $\mathrm{SL}(3, \mathbf{Z})$. *Acta Arith.*, 50(1):31–89, 1988.

[BK19] Valentin Blomer and Rizwanur Khan. Twisted moments of L-functions and spectral reciprocity. *Duke Math. J.*, 168(6):1109–1177, 2019.

[BLM19] Valentin Blomer, Xiaoqing Li, and Stephen D. Miller. A spectral reciprocity formula and non-vanishing for L-functions on $\mathrm{GL}(4)\times\mathrm{GL}(2)$. *J. Number Theory*, 205:1–43, 2019.

[Blo13] Valentin Blomer. Applications of the Kuznetsov formula on $GL(3)$. *Invent. Math.*, 194(3):673–729, 2013.

[Bum84] Daniel Bump. *Automorphic forms on* $\mathrm{GL}(3, \mathbf{R})$, volume 1083 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1984.

[But13] Jack Buttcane. On sums of $\mathrm{SL}(3, \mathbb{Z})$ Kloosterman sums. *Ramanujan J.*, 32(3):371–419, 2013.

[Fri87] Solomon Friedberg. Poincaré series for $\mathrm{GL}(n)$: Fourier expansion, Kloosterman sums, and algebreo-geometric estimates. *Math. Z.*, 196(2):165–188, 1987.

[Gar05] Paul Garrett. Buildings, Bruhat decompositions, unramified principal series, 2005.

[GK13] Dorian Goldfeld and Alex Kontorovich. On the GL(3) Kuznetsov formula with applications to symmetry types of families of L-functions. In *Automorphic representations and L-functions*, volume 22 of *Tata Inst. Fundam. Res. Stud. Math.*, pages 263–310. Tata Inst. Fund. Res., Mumbai, 2013.

[HL19] G. H. Hardy and J. E. Littlewood. A new solution of Waring's problem. *Quart. J.*, 48:272–293, 1919.

[HR18] G. H. Hardy and S. Ramanujan. Asymptotic Formulaae in Combinatory Analysis. *Proc. London Math. Soc. (2)*, 17:75–115, 1918.

[Klo27] H. D. Kloosterman. On the representation of numbers in the form $ax^2+by^2+cz^2+dt^2$. *Acta Math.*, 49(3-4):407–464, 1927.

[KN20] Eren Mehmet Kıral and Maki Nakasuji. Parametrization of Kloosterman sets and SL_3-Kloosterman sums. *arXiv preprint arXiv:2001.01936*, 2020.

[Li11] Xiaoqing Li. Bounds for $\mathrm{GL}(3) \times \mathrm{GL}(2)$ L-functions and GL(3) L-functions. *Ann. of Math. (2)*, 173(1):301–336, 2011.

[LY12] Xiaoqing Li and Matthew P. Young. The L^2 restriction norm of a GL_3 Maass form. *Compos. Math.*, 148(3):675–717, 2012.

[Mot93] Yoichi Motohashi. An explicit formula for the fourth power mean of the Riemann zeta-function. *Acta Math.*, 170(2):181–220, 1993.

[Mot97] Yoichi Motohashi. *Spectral theory of the Riemann zeta-function*, volume 127 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1997.

[Pet15] Ian N. Petrow. A twisted Motohashi formula and Weyl-subconvexity for L-functions of weight two cusp forms. *Math. Ann.*, 363(1-2):175–216, 2015.

[Ste87] Glenn Stevens. Poincaré series on $\mathrm{GL}(r)$ and Kloostermann sums. *Math. Ann.*, 277(1):25–51, 1987.

[You16] Matthew P. Young. Bilinear forms with GL_3 Kloosterman sums and the spectral large sieve. *Int. Math. Res. Not. IMRN*, (21):6453–6492, 2016.

[Zac19] Raphaël Zacharias. Periods and Reciprocity I. *International Mathematics Research Notices*, 06 2019. rnz100.

Department of Information and Communication Science, Faculty of Science,, Sophia University,, 7-1 Kioi-cho, Chiyoda-ku, Tokyo, 102-8554, Japan

Email address: erenmehmetkiral@protonmail.com

Email address: nakasuji@sophia.ac.jp